

WHAT IS CLAIMED IS:

1. A computer implemented method for predicting the structure of a membrane-bound protein having a plurality of  $\alpha$ -helical regions, comprising:

    providing an amino acid sequence for the membrane-bound protein;

    using the amino acid sequence to identify one or more transmembrane regions of the membrane-bound protein;

    constructing a set of helices for the transmembrane regions and optimizing a helix bundle configuration for the set of helices using a first molecular dynamics simulation;

    constructing a plurality of inter-helical loops to generate a full-atom model of the membrane-bound protein;

    optimizing the full-atom model using a second molecular dynamics simulation; and

    outputting a predicted structure for the transmembrane protein based on the second optimization.

2. A computational model of the structure of a transmembrane protein having a plurality of  $\alpha$ -helical regions, the computational model comprising:

    a computer-readable memory storing data describing an optimized predicted three-dimensional structure for the

transmembrane protein, the optimized predicted structure being generated according to the method of claim 1.

3. The method of claim 1, wherein:

constructing the set of helices for the transmembrane regions includes constructing a set of canonical helices corresponding to the transmembrane regions, calculating a minimum-energy configuration for each of the canonical helices, optimizing each of the canonical helices, assembling a helix bundle including each of the set of helices, and calculating a minimum-energy configuration for the helix bundle in a lipid bilayer.

4. A computational method for modeling the structure of a transmembrane protein having a plurality of  $\alpha$ -helical regions, the method comprising:

providing amino acid sequence information and sequence alignment information for a transmembrane protein having a plurality of  $\alpha$ -helical regions;

using the amino acid sequence information and the sequence alignment information to predict a set of transmembrane segments of the transmembrane protein;

constructing canonical helices for the predicted transmembrane segments and optimizing the canonical helices using a first molecular dynamics simulation;

combining the optimized helices based on the sequence alignment information to form a helix bundle, and assembling the helix bundle with a lipid bilayer to form a system helix bundle;

optimizing the structure of the system helix bundle using a second molecular dynamics simulation;

adding inter-helical loops to the system helix bundle to form a full atom model;

optimizing the full atom model using a third molecular dynamics simulation; and

outputting a predicted structure for the transmembrane protein based on the third optimization.

5. The method of claim 4, wherein:

the transmembrane protein is a G-protein coupled receptor.

6. The method of claim 4, wherein:

an energy minimum is calculated for each of the canonical helices before forming the helix bundle.

7. The method of claim 4, further comprising:

determining the periodicity of hydrophobic residues identified in the amino acid sequence information; and identifying a plurality of lipid-accessible residues based at least in part on the identified periodicity.

8. The method of claim 4, wherein:

combining the helices based on the sequence information to form a helix bundle includes orienting the helix axes according to the 7.5 Å electron density map for rhodopsin.

9. The method of claim 7, wherein:

combining the helices based on the sequence information to form a helix bundle includes orienting the identified lipid-accessible residues to face the outside of the helix bundle.

10. The method of claim 4, wherein:

the first molecular dynamics simulation is a torsional molecular dynamics simulation.

11. The method of claim 4, wherein:

the second molecular dynamics simulation is a rigid body molecular dynamics simulation.

12. The method of claim 11, wherein:

the first molecular dynamics simulation is a Newton-Euler Inverse Mass Operator dynamics simulation.

13. The method of claim 4, wherein:

the third molecular dynamics simulation is a mixed mode molecular dynamics simulation.

14. The method of claim 4, wherein:

at least the third molecular dynamics simulation includes a solvent approximation.

15. The method of claim 14, wherein:

the solvent approximation is a continuum solvation model.

16. The method of claim 15, wherein:

the solvent approximation includes the Surface Generalized Born model.

17. The method of claim 15, wherein:

the solvent approximation includes the Poisson-Boltzmann description.

18. The method of claim 14, wherein:

the solvent approximation is an empirical approximation comprising estimating solvation free energy as a function of solvent accessible protein surface area.

19. The method of claim 4, wherein:

the predicted structure is generated by performing the third molecular dynamics simulation for a time in the range from about 100ps to about 1 ns.

20. A computational model of the structure of a transmembrane protein having a plurality of  $\alpha$ -helical regions, the computational model comprising:

a computer-readable data storage medium storing data describing an optimized predicted three-dimensional structure for the transmembrane protein, the optimized predicted structure being generated according to the method of claim 4.

21. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S6.

22. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S18.

23. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S19.

24. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S25.

25. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S46.

26. The computational model of claim 20, wherein:

the transmembrane protein is olfactory receptor S50.

27. A computer program product on a computer-readable medium for predicting the structure of a membrane-bound protein having a plurality of  $\alpha$ -helical regions, the computer program product comprising instructions operable to cause a programmable processor to:

provide an amino acid sequence for the membrane-bound protein;

use the amino acid sequence to identify one or more transmembrane regions of the membrane-bound protein;

construct a set of helices for the transmembrane regions and optimize a helix bundle configuration for the set of helices using a first molecular dynamics simulation;

construct a plurality of inter-helical loops to generate a full-atom model of the membrane-bound protein;

optimize the full-atom model using a second molecular dynamics simulation;

output a predicted structure for the transmembrane protein based on the second optimization.

28. A computer program product on a computer-readable medium for predicting the structure of a G-protein coupled receptor having a plurality of  $\alpha$ -helical regions, the computer program product comprising instructions operable to cause a programmable processor to:

provide amino acid sequence information and sequence alignment information for a G-protein coupled receptor;

use the amino acid sequence information and the sequence alignment information to predict a set of transmembrane segments of the G-protein coupled receptor;

construct canonical helices for the predicted transmembrane segments and optimize the canonical helices using a first molecular dynamics simulation;

combine the optimized helices based on the sequence alignment information to form a helix bundle and assemble the helix bundle with a lipid bilayer to form a system helix bundle;

optimize the structure of the system helix bundle using a second molecular dynamics simulation;

add inter-helical loops to the system helix bundle to form a full atom model;

optimize the full atom model using a third molecular dynamics simulation; and

output a predicted structure for the G-protein coupled receptor based on the second optimization.

29. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S6, the data describing the predicted structure including structure coordinates for the amino acids of olfactory receptor S6 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 2 of less than or equal to about 2.0 angstroms.

30. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S18, the data describing the predicted structure including

structure coordinates for the amino acids of olfactory receptor S18 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 3 of less than or equal to about 2.0 angstroms.

31. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S19, the data describing the predicted structure including structure coordinates for the amino acids of olfactory receptor S19 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 4 of less than or equal to about 2.0 angstroms.

32. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S25, the data describing the predicted structure including structure coordinates for the amino acids of olfactory receptor S25 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 5 of less than or equal to about 2.0 angstroms.

33. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S46, the data describing the predicted structure including structure coordinates for the amino acids of olfactory receptor S46 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 6 of less than or equal to about 2.0 angstroms.

34. A computational model of the structure of a G-protein coupled receptor, the computational model comprising:

a computer-readable data storage medium storing data describing a three-dimensional structure for olfactory receptor S50, the data describing the predicted structure including structure coordinates for the amino acids of olfactory receptor S50 having a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids as set out in Table 7 of less than or equal to about 2.0 angstroms.